REMARKS

Claims 1-5, 7-14 and 16-20, of which Claims 1, 10, 19 and 20 are the independent claims, are pending in the application. Claims 1-5, 7-9 and 19 (the pending method claims) have been rejected under 35 U.S.C. §101. Claims 1-5, 7-14, 16-20 have been rejected under 35 U.S.C. §112, first paragraph. Claim 1-5, 7, 8, 10-14 and 16-20 have been rejected under 35 U.S.C. §102(b) and Claims 1-5, 7-14, and 16-20 have been rejected under 35 U.S.C. §103(a).

Applicants respond as follows.

Claim Amendments

Claims 1 and 19 have been amended to recite a step of analyzing the subject genome sequence by using the uniform representation (e.g., as input) in the analysis. The support for this amendment is found at least on Specification page 10, line 6 - page 11, line 12 as originally filed.

Claims 10 and 20 have been amended to incorporate the subject matter of Claims 17 and 18.

Claims 17 and 18 have now been cancelled.

New Claims 21 - 25 have been added to further define the method of the invention. Support for Claim 21 is found throughout the Specification as originally filed and, for example, on page 7, lines 4-6; page 9, lines 11-19; and page 10, lines 3-7. Support for Claim 22 is found at least on Specification page 6, lines 27 - 28 as originally filed. Support for Claim 23 is found at least on Specification page 9, lines 3 - 6 as originally filed. Support for Claim 24 is found at least on Specification page 8, lines 12 - 18 and on page 9, lines 6 - 9 as originally filed. Support for Claim 25 is found at least on Specification page 9, lines 6 - 9 as originally filed.

New Claims 26 - 30 have been added to further define the apparatus of the invention. Support for Claim 26 is found throughout the Specification as originally filed, for example, on page 7, lines 4-20 and page 9, lines 11-19, in Claim 10 as originally filed, and page 10, line 5 - page 11, line 12. Support for Claim 27 is found at least on Specification page 6, lines 27-28 and page 7, lines 3-10 as originally filed. Support for Claim 28 is found at least on Specification page 9, lines 3-6 as originally filed. Support for Claim 29 is found at least on Specification page 8, lines 12-18 and on page 9, lines 6-9 as originally filed. Support for Claim 30 is found at least on Specification page 9, lines 6-9 as originally filed.

No new matter is introduced by this Amendment.

Rejection of Claims 1 - 5, 7 - 9 and 19 under 35 U.S.C. §101

Claims 1 - 5, 7 - 9 and 19 have been rejected under 35 U.S.C. §101 as being directed to non-statutory subject matter. The Office Action states that the claims are drawn to methods of manipulating data which do not produce a concrete, tangible and useful result.

Independent method Claims 1 and 19 have now been amended. As now amended, Claims 1 is drawn to a method for analyzing a physical object, namely a subject genome sequence. The method operates by accepting two inputs, namely a set of known biological fragments and a subject genome sequence, and by producing an output: a classification, clustering or indexing of the subject genome sequence. This classification, clustering or indexing is of unquestionable value: knowledge of structure and/or function of a gene or a gene product opens innumerable opportunities that range from drug design to function prediction to structure prediction to genome annotation and indexing (pages 1 and 2 of the Specification). The central role that methods of molecular structure/function determination (*i.e.* assignment of a molecule into a structural and/or functional class) such as X-ray crystallography and sequence alignment program BLAST (Altshul *et al.*, 1990, Basic Local Alignment Search Tool, JMB 215:403-410) have played and continue to play in biomedical arts undeniably attests to usefulness and "real world value" of these methods. Among "real world" examples of applications of these methods are cancer gene therapy, insulin therapy for diabetes and HIV protease inhibitors. Thus the invention of Claim 1 is something more than an abstract idea or a concept: the steps of Claim 1 generate a new "concrete, tangible and useful result" (i.e., an analyzed, namely classified, clustered and/or indexed, subject genome sequence) within the meaning of the *State Street* (See *State Street*, 149 F.3d at 1373, 47 USPQ2d at 1601-02).

The same argument applies to Claim 19, as now amended, and directed to a method for analyzing a subject protein sequence.

It is respectfully requested that the §101 rejection be withdrawn.

Rejection of Claims 1 - 5, 7 -14 and 16 - 20 Under 35 U.S.C. §112

Claims 1 - 5, 7 -14 and 16 - 20 have been rejected under 35 U.S.C. §112. It is Applicants' understanding that the Examiner reiterates her reasons in support of this rejection as set forth in the Office Action mailed on October 1, 2002. The Examiner states that the methods and apparati of the present invention are not described with details sufficient to enable one skilled in the art to practice the present invention without undue experimentation. The Examiner also states that the Applicant sets forth that the set of known biological fragments to be used is an annotated protein sequence database, that the Specification does not provide any other types of sets of biological fragments that can be used and that the claims do not recite the need for such an annotated database. The Examiner further states that, while an annotated protein sequence database appears to be a critical element, it is missing from all the claims.

The Applicants respectfully disagree with the Examiner. The present invention is not directed to a new or specific database, nor a method for creating a database, but involves use of the content of existing databases well-known in the art. The Specification clearly states on page 8, line 9, that the method "creates *or* obtains a comparison database". (Emphasis added.) The Specification further provides examples of publically available databases suitable for practicing the present invention: the BLOCKS database (Steven Henikoff and Jorja G. Henikoff, "Automated assembly of protein blocks for database searching," *Nucleic Acids Research, 19:23*, pp. 6565-6572 (1991)), Emotif (http://dna.stanford.edu/emotif/), and PRINTs (http://bioinf.man.ac.uk.dbbrowser/ PRINTS/). A skilled artisan can access these databases and thus provide a predefined set of known biological fragments called for by the claimed method of the invention. Likewise, these databases, stored either on a remote host or downloaded onto a local memory device, provide a data store of respective representations of a set of a predefined number of known biological fragments called for by the apparatus claims. The Applicants submit that three working examples certainly constitute enough guidance to avoid undue experimentation.

The Applicants further submit, that the Specification provides sufficient guidance for one skilled in the art of biological sequence analysis to create a database suitable for practicing the invention. Referring to page 8, lines 25 - 29, the Specification teaches that short sequences (line 27) stored in a database labeled according to structure (lines 25 - 26) can be multiply aligned

(line 27) using publically available software (two examples of such software are given) and that the statistics can be collected about these short sequences. Specification page 8, lines 12 - 18, also discloses a preferred embodiment of a representation of a set of biological structures as a matrix of probabilities. A skilled artisan would appreciate that such a matrix is obtained by statistical analysis of multiple aligned sequences. Furthermore, it is common knowledge among practitioners of the art of sequence analysis that sequences in a database, particularly protein sequence databases, are labeled or annotated to indicate structural or functional domains. In view of the above, the Applicants submit that the present disclosure provides sufficient guidance for one ordinarily skilled in the art to create a set of biological fragments (*e.g.*, domains), each fragment having a representation (*e.g.*, probability matrix).

With regard to apparatus claims, the Examiner states that significant database information is lacking from the claims as well as a description of how the processor acts on that information to provide some results.

The Applicants respectfully disagree. In view of the arguments presented above, the Applicants submit that the Specification fully describes "a data store of representations of a predefined number of known biological sequences" called for by base Claim 10 and therefore enables one ordinarily skilled in the art to practice the present invention without undue experimentation. Furthermore, the Specification discloses two working examples of "a comparison routine executed by a digital processor having access to the data store": a probability of the subject genome sequence being generated by the known biological sequence and a counting of a number of occurrences of the known biological sequence found in the subject genome sequence (Specification page 9, lines 3 - 9). Claims 17 and 18 are drawn to these specific examples of comparison routines. However, in the interest of facilitating prosecution, Claim 10 has now been amended to recite the subject matter of Claims 17 and 18, and Claims 17-18 are now canceled.

According to the foregoing, base method and apparatus Claims 1, 10, 19 and 20 and claims dependent thereon (i.e., Claims 2-5, 7-14 and 16) are believed to contain subject matter which is described in the Specification in such a way as to enable one skilled in the art to make and/or use the invention. As such, it is believed that Claims 1-5, 7-14, 16 and 19-20 as now amended comply with the enablement requirement of § 112.

Reconsideration and withdrawal of the §112 rejection is respectfully requested.

Rejection of Claims 1 - 5, 7, 8, 10 - 14 and 16 - 20 Under 35 U.S.C. 102(b) over Akutsu (1994) and Akutsu *et al.* (1997)

The Office Action states that Akutsu, T., "Substructure Search and Alignment for Three-dimensional Protein Structure", Joho Shori Gakkai Kenkyu Hokoku, (1994) vol. 94 no. 82 (AL.41), pp. 1 - 8 (hereinafter, Akutsu (1994)) discloses methods of generating hash vectors from a fixed set of biological fragments from a protein structure or structure database and that these hash vectors are manipulated as to the number of occurrences, relatedness to other sequences etc. The Office Action states that this appears to meet the limitations of Claim 1, and that the computer system that perform these methods meets the limitations of the apparatus claims.

The Office Action states that Akutsu *et al.*, "Rapid Protein Fragment Search Using Hash Functions Based on the Fourier Transform", CABIOS (1997) Vol. 13:4, pages 357 - 364 (hereinafter, Akutsu *et al.* (1997)) discloses a variation of methods of Akutsu (1994) that utilizes hash vectors from fixed length fragments from biological sequences. The Office Action thus reaches the conclusion that Akutsu *et al.* (1997) anticipates Claim 1 and the apparatus claims.

The Applicants believe that the references of Akutsu (1994) and Akutsu *et al.* (1997) have been erroneously cited against the present invention.

Akutsu (1994) and Akutsu *et al.* (1997) relate to the art of computational geometry, more specifically to geometric hashing. By way of introduction, geometric hashing is a technique of finding common subfigures, invariant under rotation, translation and scale, in two or more (usually, three) dimensions. When applied to molecular structure analysis, geometric hashing is a technique for matching a three-dimensional structure of a target molecule (or a collection of such) against a set of one or more models (*e.g.*, a database) known in advance. Accordingly, Akutsu (1994) and Akutsu *et al.* (1997) teach a method for searching for similar *three-dimensional* fragments among the entries of a protein structure database. (See Akutsu (1994), Abstract and Akutsu *et al.* (1997), Abstract and Introduction, p. 357, first paragraph.) The Akutsu method utilizes a geometric hashing technique whereby a low frequency Fourier spectrum of distances between each Cα carbon of a structure or a fragment of a structure under consideration and the averaged center of all Cα carbons of the a structure under consideration

("centroid") forms a vector. These vectors are then compared using any of the standard techniques (Akitsu *et al.* (1997), page 359). Additionally, Akutsu (1994) discloses a "least-square hashing" method (Akutsu (1994), pages 2-3) whereby a root-mean-square-like function *d* is minimized for similar structures. Furthermore, "fixed length fragments from biological sequences" mentioned by the Examiner refers to the *input* of the Akutsu's algorithm (see Akutsu *et al.* (1997), Abstract), rather than output, as in the method of the present invention.

The Applicants submit that neither method nor apparatus claims of the present invention are anticipated by either Akutsu (1994) or Akutsu *et al.* (1997). There is no teaching in either reference of a set of known biological fragments, the set being of a fixed number of said known biological fragments, each known biological fragment in the set having a respective representation as recited in the first paragraph of each of base Claims 1, 10 and 19. As the Applicants pointed out above, examples of the representations include text strings (sequences) and probability matrices, certainly not sets of three-dimensional coordinates as in the Akutsu method.

There is no teaching in either reference of quantitatively determining a score of each biological fragment in the set with respect to (e.g., as compared against) the subject genome sequence, said scores forming a feature vector having a length equal to the predefined number of known biological sequences. Such is recited in lines 6-16 of base Claim 1 as now amended, lines 4-8 and 12-14 of base Claim 10 as now amended, subparagraphs (b) and (c) of base Claim 19 as now amended, lines 5-8 of base Claim 20, subparagraphs (c) and (d) of new base Claim 21 and subparagraph (c) of new base Claim 26. Rather, the Akutsu method forms a hash vector for each query and only then compares the query vector to the vectors corresponding to the members of the database. Additionally, the number of components ("length") of an Akutsu's hash vector is determined by the number of three-dimensional points comprising a query structure, rather than by a number of structures in the database as is specified for the claimed present invention.

Finally, neither reference teaches a comparison routine (or scoring routine that uses comparisons), an element of the apparatus claims (Claims 10-14, 16, 20, 26, 28 and 30) of the present invention, said comparison routine comparing each known biological sequence to a subject genome sequence and generating a score, said scores forming a vector having a length equal to the predefined number of known biological sequences, wherein the generated score is

either a probability of the subject genome sequence being generated by the known biological sequence or a counting of a number of occurrences of the known biological sequence found in the subject genome sequence. See the second subparagraph of base Claim 10 and last lines of Claim 20 as now amended reciting the foregoing language. Also see similar terms in new apparatus Claims 26, 28 and 30.

Thus, for the foregoing reasons, the cited art does not imply, suggest or in any way anticipate the present invention as claimed in base Claims 1, 10, 19-21 and 26, and by virtue of their respective dependencies, dependent Claims 2-5, 7-8, 11-14, 16, 22-25 and 27-30.

Reconsideration and withdrawal of the §102 rejections is respectfully requested.

Rejection of Claims 1 - 5, 7 - 14 and 16 - 20 Under 35 U.S.C. 103(a) over Berry *et al.*

Claims 1 - 5, 7 - 14 and 16 - 20 have been rejected under 35 U.S.C. 103(a) over Berry *et al.*, "Matrices, Vector Spaces and Information Retrieval", SIAM Rev. 41:2 335 - 362 (1999) (hereinafter "Berry *et al.*"). The Office Action states that Berry *et al.* appears to disclose the same mathematical concepts as the invention, and teach the application of the concepts to words, text documents etc., all in digital databases. The Office Action further states that these are the steps required by the rejected claims. Additionally, the Office Action states that the difference between the cited reference and the claimed invention is in the nature of the text information.

The Applicants disagree with the Office Action's assertion that Berry *et al.* discloses the steps required by the rejected claims.

Berry *et al.* describes two information retrieval and comparison methods that utilize simple linear algebra operations. One such method is "Term-Term Comparison", disclosed on pages 352 - 354. The other method is "Query Matching", disclosed on pages 340 - 342.

Term-Term Comparison is a process that computes correlations (frequencies of co-occurrence) between the "terms" (words) of a database. The input of this method is a "term-by-document" (frequency-of-occurrence) matrix *G*, shown in FIG. 7.1 on page 353, that describes frequencies with which seven terms occur in five documents. Correlations are computed as simple scalar products of all pairs of term vectors (row vectors of *G*) in a Eucledian vector space according to Eq. (7.1). The output is a symmetric matrix *C* shown in FIG. 7.1, where each entry C_{ij} gives a correlation coefficient between term vectors *i* and *j*. A coefficient of 1 means that

these terms always occur together, while a coefficient of 0 indicates that the terms never occur together. Furthermore, since the correlation coefficients are scalar products of vectors, geometric separation (spacial grouping) of term vectors may be inferred. According to Berry *et al.*, the process of grouping terms according to their related content in this way is known as clustering (page 353, last full paragraph). Clearly, neither a process nor an apparatus, whose sole input is a "frequency of occurrence" matrix and whose sole output is a matrix of correlation coefficients between the row vectors (or even geometrical clusters of row vectors) is claimed by the present invention. The Applicants submit that in this regard, no claim of the present invention is made obvious by the Term-Term Comparison method of Berry *et al.*

Query Matching is a method of retrieval of documents that best match a query. There are two inputs. The first input is a frequency-of-occurrence "term-by-document" matrix A shown in FIG. 2.2. The second input is a query vector q with components being either 0 or 1, depending on whether a term is present or absent in the query. The method returns a column vector that is a product of matrix A and vector q. The components of this vector are correlation coefficients which are scalar product of term vectors (row vectors of A) and vector q in a Eucledian vector space. Based on these correlation coefficients, the documents are classified as "relevant" or "not relevant".

It appears that the Examiner is equating:

matrix A with a probabilistic template representation of any one biological fragment in a set,

a "word" with a nucleotide or an amino acid,

a "document" with a position in a sequence,

query vector q with a subject genome sequence; and

the qA product with a score.

The Applicants submit that the steps of the Query Matching method do not meet the limitations of the claims of the present invention. Superficial resemblance of matrix A to a probabilistic template representation of biological fragments, a query vector q to a subject genome sequence, and a product of q and A to a score does not stand under a closer examination.

Firstly, the present invention calls for a set of biological fragments (or biological sequences). See base Claims 1, 10, 19-21 and 26. Each fragment in a set is represented by a respective representation which, in one embodiment, is a respective probabilistic template. Thus, in one embodiment, the invention as claimed requires a set of probabilistic templates, or matrices. Berry *et al.* does not appear to contemplate any Query Matching method that uses more than one frequency-of-occurrence matrix.

Secondly, the method of the present invention calls for quantitative determination of a score of each biological fragment in the set against the subject genome sequence and for forming a feature vector of the subject genome sequence, said feature vector being a sequence of scores of each biological fragment in the set. It is clear from both the Specification and the claims (base Claims 1, 10, 19-21 and 26 now amended or presented) that a score is a number (a scalar). Multiplying a query vector q by matrix A produces a vector in Berry *et al.* There is no guidance in Berry *et al.* as to how to use the product of q and A to produce a scalar score.

Thirdly, Berry *et al.* certainly do not teach forming a feature vector, said vector being a sequence of the scores of each biological fragment, and the number of components ("length") of the feature vector being equal to the number of fragments in a set. More than one fragment in a set implies more than one matrix A. As mentioned above, Berry *et al.* do not contemplate existence of different "term-by-word" matrices. Furthermore, Berry *et al.* misses altogether the claimed formed vector having a length equal to the fixed number of fragments in the set. That is, all feature vectors formed, according to the claimed method of the present invention, have the same length. As a result, the claimed formed vectors, for each of many subject genome sequences of varying length, provide a uniform (same fixed number of components) representation of the subject genome sequences. The cited art does not imply or suggest such transformation from varying length query vectors q to uniform (same fixed number of components) representations as in the claimed invention. See base Claim 1, lines 15-18, base Claim 10, lines 7-12, base Claim 19, subparagraph (c) and base Claim 20, lines 7-10.

Lastly, the subject genome sequence of the present invention can be of arbitrary length, whereas a query vector q of Berry *et al.* is limited to the number of components equal to the number of terms. Additionally, the query vector q by necessity must contain components that are

either 0's or 1's, whereas a subject genome sequence of the present invention may have either 5 (for (deoxy)ribonucleotides) or 20 (for amino acids) values for each component.

Additionally, it is noted that Berry *et al.* describe the methods that utilize vector algebra. Accordingly, one embodiment of the present invention wherein biological fragments are represented by text strings and scoring includes counting the number of times each string is found within the subject genome sequence is neither described nor suggested by Berry *et al.*

As to the apparatus claims, none of the computer systems employed by Berry *et al.* meet the limitations of Claims 10-14, 16, 20 and 26 - 30; for the reasons presented above it is clear that neither comparison routine of Claims 10 and 20 nor a scoring routine of Claim 26 is taught or suggested by Berry *et al.*

For the foregoing reasons, it is believed that the present invention as recited in base Claims 1, 10, 19-21 and 26 and (by virtue of their dependencies) as inherently in Claims 2-5, 7-14, 16, 19, 22-25 and 27-30 is not made obvious by the cited or prior art.

Reconsideration and withdrawal of the §103 rejection is respectfully requested.

CONCLUSION

In view of the above amendments and remarks, it is believed that all now pending claims (Claims 1-5, 7-14, 16 and 19-30) are in condition for allowance, and it is respectfully requested that the application be passed to issue. If the Examiner feels that a telephone conference would expedite prosecution of this case, the Examiner is invited to call the undersigned at (978) 341-0036.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By _____

Mary Lou Wakimura
Registration No. 31,804
Telephone: (978) 341-0036
Facsimile: (978) 341-0136

Concord, MA 01742-9133
Dated: 8/5/04